

Security and Privacy in Cloud Computing

Ragib Hasan
Johns Hopkins University
en.600.412 Spring 2010

Lecture 7
03/29/2010

Provenance

- **Provenance:** from Latin *provenire* 'come from', defined as
  - "(i) *the fact of coming from some particular source or quarter; origin, derivation.*
  - *(ii) the history or pedigree of a work of art, manuscript, rare book, etc.; a record of the ultimate derivation and passage of an item through its various owners*" (Oxford English Dictionary)
- In other words, **Who** owned it, **what** was done to it, **how** was it transferred …
- Widely used in arts, archives, and archeology, called the **Fundamental Principle of Archival**

http://moma.org/collection/provenance/items/644.67.html

L'artiste et son modèle (1928), at Museum of Modern Art

Data Provenance

- Definition*
  - Description of the **origins** of data and the **process** by which it arrived at the database. [Buneman et al.]
  - Information describing materials and **transformations** applied to derive the data. [Lanter]
  - Metadata recording the **process of experiment workflows**, annotations, and notes about experiments. [Greenwood]
  - Information that helps determine the **derivation history** of a data product, starting from its original sources. [Simmhan et al.]

*Simmhan et al. A Survey of Provenance in E-Science. SIGMOD Record, 2005.

Forensics and Provenance in Clouds

- Cloud provenance can be
  - **Data provenance**: Who created, modified, deleted data stored in a cloud (external entities change data)
  - **Process provenance**: What happened to data once it was inside the cloud (internal entities change data)
- Cloud provenance should give a **record** of who accessed the data at different times
- Auditors should be able to **trace** an entry (and associated modification) back to the creator

Privacy questions

- Should the cloud provider know the identity of cloud users?
- Should cloud users know the identity of other users in the same group?

The “Bread and Butter” paper

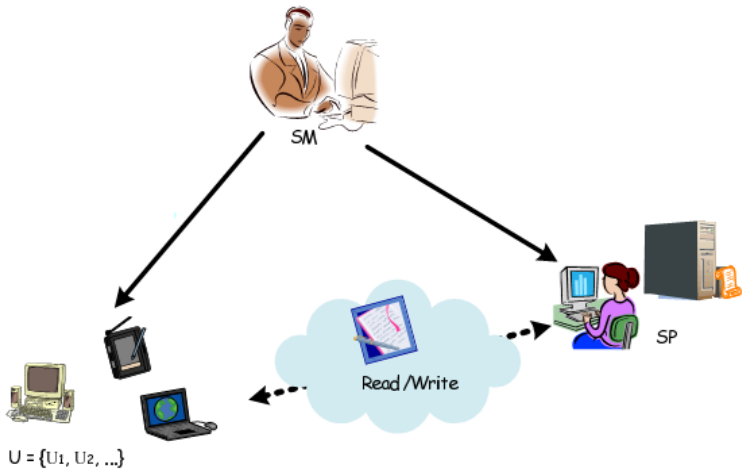
Problem

- To preserve user privacy and allow anonymous authentication/access in a cloud
- To determine authorship of data, i.e., to bind data versions to user identities in a cloud

Lu et al., Secure Provenance: The Essential Bread and Butter of Data Forensics in Cloud Computing, AsiaCCS 2010

Threat Model

- Who are the key players?
  - Users
  - SM
  - SP
- Who trusts who?
  - Users: trust the SM, but not the SP
  - SP: Trust SM
  - SM: ?
- What attacks can happen?



System Model

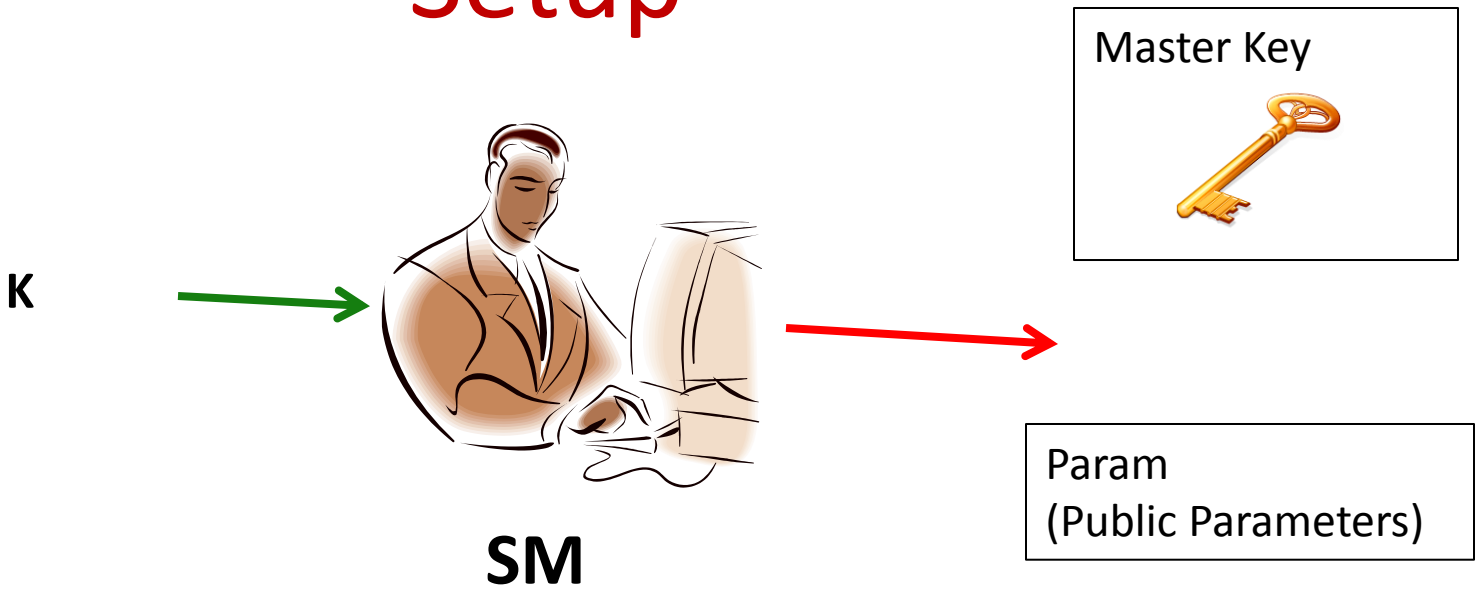
- **SM**: Manages the whole system(?), registers cloud users and providers, issues keys
- **SP**: Cloud service provider, manages access to cloud resources
- **Users**: A user is part of a group of authorized principals who can access group resources

Secure provenance (according to the paper)

By secure provenance, the authors imply

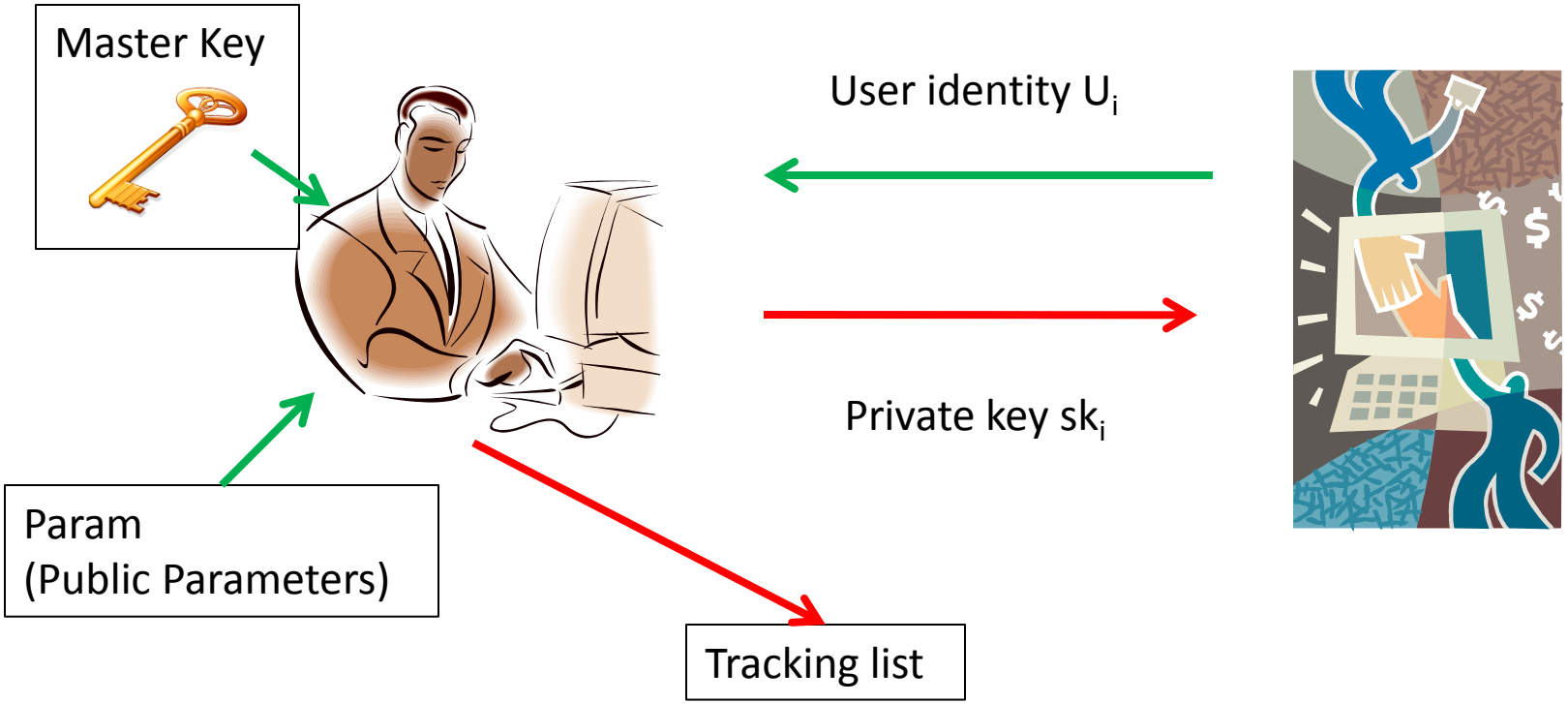
- Users **can anonymously authenticate** themselves as part of authorized users/groups to the cloud provider
- Users can **anonymously access** and modify resources
- **Encrypted data** stored by a user can be decrypted by other users from the same group
- If necessary, the SM can **trace** a data item to the user who created it

Setup



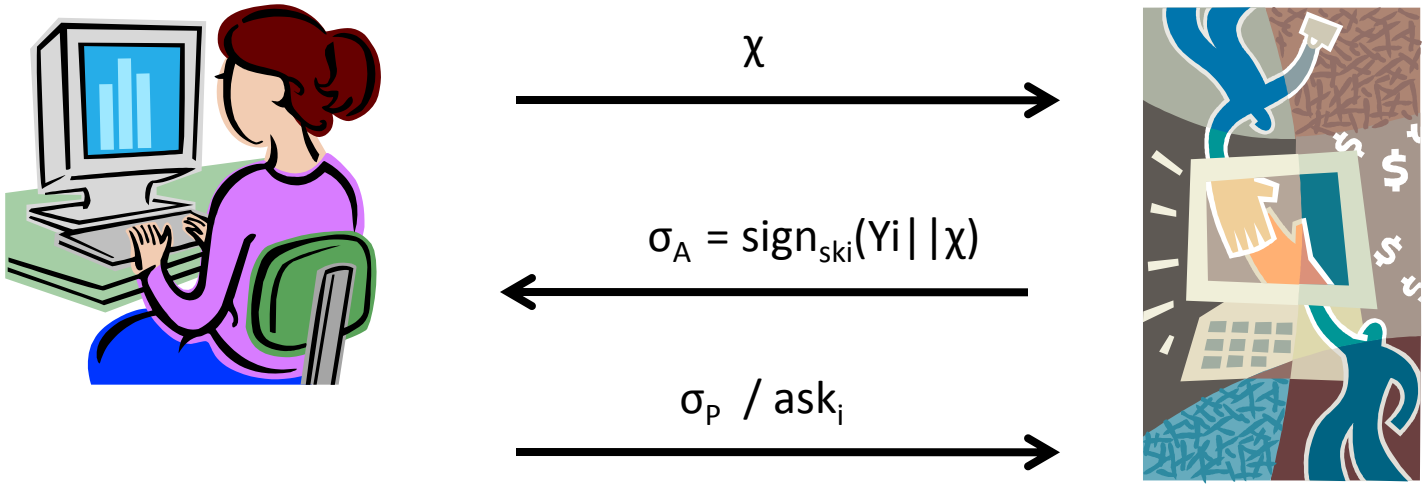
- Inputs: Security parameter k
- Output: Master key, public parameters

User/provider registration



- Inputs: Master key, public parameters, user identity
- Outputs: Private key, entry in tracking list

User-cloud interaction (1)



User anonymously authenticate herself to the cloud

Cloud provider can check that the signature was made with a key issued by the SM

User-cloud interaction (2)

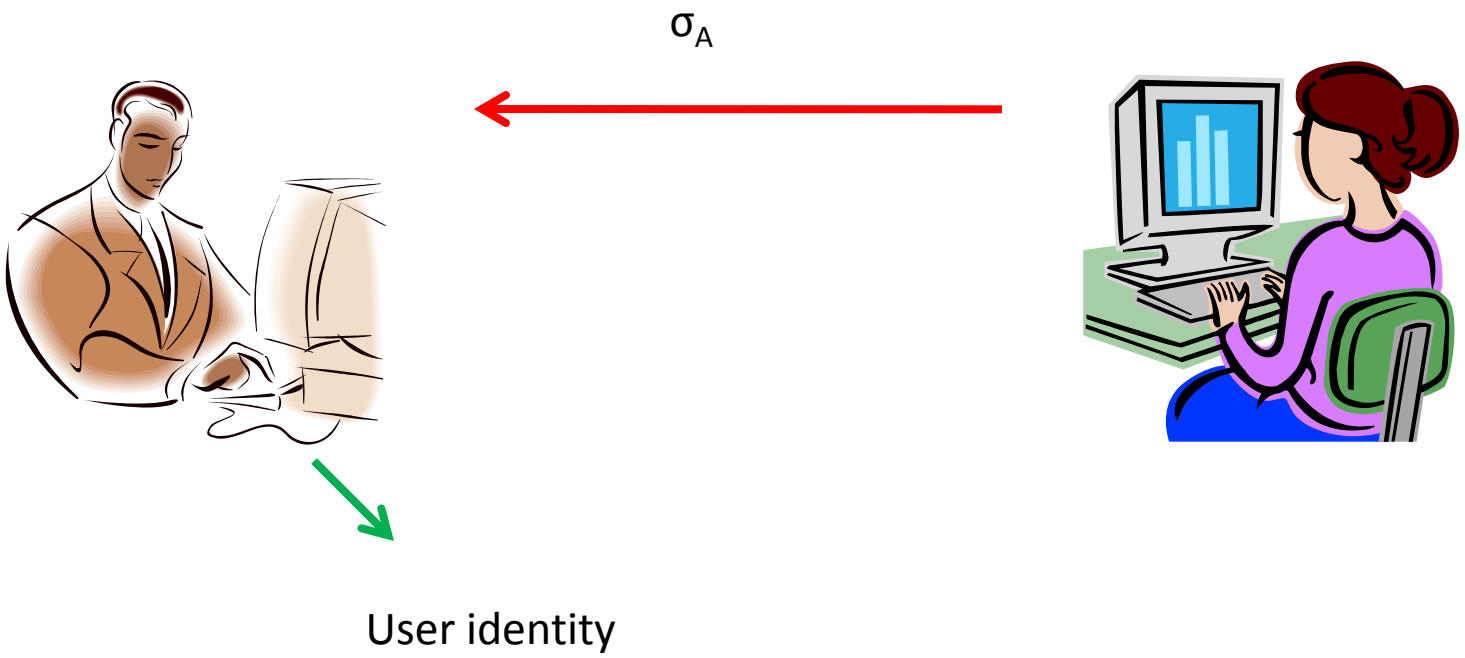
EncryptedData: C = encrypt(M)

$Sig = sign_{aski}(C)$

Store C and σ_A

Provider stores Signatures and authentication information during each access

Identifying authorship



Confidentiality preservation

- Each user gets a different authorized group user access key
- Any group user access key can be used to decrypt a ciphertext created by other users in the same group

Discussion

Suppose Amazon S3 implements such a model. What will be the advantages, and what will be the disadvantages?

What about other provenance in computation clouds?

If the data is being manipulated by processes running in the cloud, how will the problem change?

Further Reading

Ragib Hasan, Radu Sion, and Marianne Winslett, Protecting History Forgery with Secure Provenance, ACM Transactions on Storage, December 2009